Metadata for images: emerging practice and standards

Michael Day

UKOLN: The UK Office for Library and Information Networking,
University of Bath, Bath, BA2 7AY, United Kingdom
http://www.ukoln.ac.uk/
m.day@ukoln.ac.uk

Abstract

The effective organisation and retrieval of digital images in a networked environment will depend upon the development and use of relevant metadata standards. This paper will discuss metadata formats and digital images with particular reference to the Dublin Core initiative and the standard developed by the Consortium for the Computer Interchange of Museum Information (CIMI). Issues relating to interoperability, the management of resources and digital preservation will be discussed with reference to a number of existing projects and initiatives.

1 Introduction

From prehistoric times, human communication has depended upon the creation and use of image-based information. Images have been a key component of human progress, for example, in the visual arts, architecture and geography. According to Eugene Ferguson, technological developments relating to images in the Renaissance, including the invention of printing and the use of linear perspective, had a positive effect upon the rise of modern science and engineering [1]. The invention of photography and moving-image technology in the form of film, television and video recordings has increased global dependence on communication through images.

The importance of this image material is such that many different types of organisation exist in order to create, collect and maintain collections of them. These include publicly funded bodies like art galleries, museums and libraries as well as commercial organisations like television companies and newspapers [2]. The types of image included in these collections are similarly diverse and include things like paintings, engineering diagrams, photographic prints, maps and films. Traditionally, these organisations would have kept these images in their original formats. Photographic libraries and other organisations that hold large amounts of photographic material, for example, would usually need to store monochrome photographic prints, negatives (including fragile glass plates) and colour slides [3]. Sometimes the formats kept are obsolete. For example, the Bodleian Library in Oxford holds at least one 'manuscript' in the form of a Kinora spool (an early format for recording moving images). Conversion of this to a newer format required the use of a viewer borrowed from the University of Oxford's Museum of the History of Science [4].

An increasing amount of image-based material is beginning to exist in digital form. This is partly due to the widespread digitisation of traditional image collections but an increasing amount of image-based information is being created in purely digital formats. The success of the World Wide Web as a platform for disseminating multimedia information through the Internet has also acted as an impetus for these processes. In addition, digital image-based information is also being created in vast quantities by satellites and other remote sensing devices [5]. This rapidly growing corpus of digital image-based materials, however, leaves us with serious retrieval problems. One response to this has been the development of content-based image retrieval (CBIR) systems that can recognise and retrieve information based on colour, texture or shape [6]. Alternative, but complementary, retrieval approaches build upon traditional text-based techniques developed for non-digital images. In the digital environment, this is usually referred to as metadata. The remainder of this paper will consider the role of this metadata for images and outline some of the metadata standards currently under development.

2 Metadata and its uses

Metadata is usually defined as 'data about data' but normally refers to machine-understandable structured data about data [7]. In the library community, metadata is often used to refer to the type of descriptive information contained in catalogue records; bibliographic-type information that describes resources together with selected

index headings. Metadata, therefore, acts as a basis for information retrieval but can also have other roles.

The diversity of metadata creating communities, however, results in the existence of many different metadata formats. An approximate typology based upon the underlying complexity of the various formats is outlined in Figure 1. According to this typology, there is a continuum from simple metadata like that created by Web search engines, through simple structured generic formats like Dublin Core to more complex formats which have definite structure and are specific to one particular domain or part of a larger semantic framework. Examples of these more complex formats are the MARC formats used by libraries and formats based on the Standard Generalised Markup Language (SGML).

Band One	**Band Two**	**Band Three**	
(full text indexes)	*(simple structured generic formats)*	*(more complex structure, domain specific)*	*(part of a larger semantic framework)*
Proprietary formats	Proprietary formats Dublin Core IAFA/Whois++ templates	FGDC MARC	TEI headers ISCPR EAD CIMI

Figure 1: Typology of metadata formats. Adapted from Dempsey and Heery (1998)

Different subject communities and market sectors have invested heavily in developing their own metadata formats and systems. Lorcan Dempsey and Rachel Heery have pointed out that considerable effort has been expended on developing specialist formats to ensure fitness for purpose [8].

> ... there has been investment in training and documentation to spread knowledge of the format; and, not least, systems have been developed to manipulate and provide services based on these formats.

For these reasons, metadata 'format diversity' is likely to be perpetuated over time and, indeed, new metadata formats will periodically be developed to address the perceived needs of other subject domains and communities.

The creation and maintenance of metadata will be an important part of the activity of any image archive - whether digital or non-digital. Image repositories like photographic libraries have traditionally used cataloguing techniques to store metadata about their resources. These will, where possible, record an object's creator and date of creation and may also contain information on its physical form and subject matter. Images also have their own subject indexing requirements so that although general subject classification schemes like the Dewey Decimal Classification (DDC) can be used to index images, some classification schemes have been developed specifically for image-based resources. One of the best known is ICONCLASS, an iconographic hierarchical classification system for classifying works of art developed at the University of Leiden [9]. Text-based indexing schemes include the widely used Art & Architecture Thesaurus (AAT) and the Categories for the Description of Works of Art (CDWA) [10, 11].

2.1 The Dublin Core metadata initiative

The Internet has spurred increased consideration of metadata as an aid to resource discovery. The Dublin Core (DC) is an international and interdisciplinary initiative to develop a metadata element set intended to facilitate discovery of digital resources. Dublin Core was initially conceived as a simple metadata format that could be used by the creators of resource or by Web-site maintainers. It has also, however, become a focus of interest from a variety of communities with wider interests in resource description, including librarians, archivists, museum documentation specialists and computer scientists with an interest in text markup issues [12].

International representatives of these communities have met at a series of invitational workshops, the first of which met at OCLC's headquarters at Dublin, Ohio in March 1995 which resulted in the initial proposal of an thirteen element metadata set [13]. Dublin Core elements are intended to be optional, repeatable and extensible. The requirement for extensibility was recognised from an early stage because local applications of Dublin Core would need to add additional elements according to special needs [14]. A second workshop, held at the University of Warwick in April 1996, investigated syntactical issues and discussed the development of a architectural framework that could position the Dublin Core in a wider metadata context [15]. It has always been expected that

Dublin Core will need to co-exist with metadata created in other formats and created for purposes other than resource discovery. The Warwick Framework container architecture was developed to permit the aggregation of different metadata types [16].

The third Dublin Core Workshop, the Image Metadata Workshop held in Dublin, Ohio in September 1996 and sponsored by the Coalition for Networked Information (CNI) and OCLC, specifically considered the application of the Dublin Core element set to the description of images. Dublin Core had been originally developed to describe what the initiative itself called "document-like-objects" - as a means of side-stepping what Weibel and Miller called "differences on individual notions of what constitutes a discrete object worthy of separate description" [17]. The workshop concluded that discrete and bounded images could be considered to be document-like-objects when 'fixed' - in the sense of appearing the same to all users. The workshop concluded that Dublin Core could constitute a suitable basis for the resource discovery of networked images. In addition, the workshop noted the benefits of adopting a common set of metadata elements that would support the discovery of both textual and visual resources [18].

The workshop resulted in a number of changes to the Dublin Core element set which took into account the specific requirements of images. Some of the element descriptions were changed and two new elements were added to the original thirteen. Thus the Subject and Description elements were differentiated and a new Rights Management element added. The result was fifteen core metadata elements (Figure 2). A reference definition of simple Dublin Core has been published in 1998 as Internet RFC 2413 [19].

Element name	**Brief description**
Title	Name given to the resource by the 'Creator' or 'Publisher'
Author or Creator	Person(s) or organisation(s) primarily responsible for the intellectual content of the resource
Subject and Keywords	The topic of the resource (keywords, index terms or subject classifications)
Description	Textual description of the resource
Publisher	The entity responsible for making the resource available in its current form
Other Contributor	Person(s) or organisation(s) not included under 'Creator' who have made significant (but secondary) intellectual contributions to the resource
Date	Date the resource was made available in its present form
Resource Type	The category of the resource, e.g. 'image'
Format	The data representation of the resource, e.g. 'text/html', 'JPEG', etc.
Resource Identifier	String or number used to uniquely identify the resource, e.g. an 'URI' or 'URL'
Source	Information on the work from which the resource is derived
Language	Language(s) of the intellectual content of the resource
Relation	Relationship of the resource to other resources
Coverage	Spatial locations or temporal duration characteristic of the resource
Rights Management	A link to a rights management statement

Figure 2: Dublin Core elements. Source: Adapted from Weibel and Lagoze (1997).

Further workshops helped define Dublin Core element structure and formalised additional qualifiers. Three qualifiers were defined at DC-4 in Canberra [20].

- TYPE - a sub-element name, e.g. DC.Relation.IsPartOf.
- SCHEME - specifies an externally defined scheme or standard. So a DC Subject field could be qualified as a DDC number or LCSH term, e.g. DC.Subject SCHEME="DDC21".
- LANGUAGE - the language of the element value, e.g. DC.Description LANG="en".

Currently, the easiest way to implement metadata on the Web is to embed Dublin Core in HTML META tags. The advantage of this is that metadata is stored within the document itself and could be harvested by metadata-

aware Web indexing robots [21]. Future deployment of Dublin Core on the Web, however, will possibly be implemented using the Resource Description Framework (RDF).

2.3 The Resource Description Framework

RDF has been developed by the World Wide Web Consortium (W3C), the organisation that co-ordinates standards for the Web, as an architecture for metadata on the Web. The *RDF Model and Syntax Specification* provides a data model for describing resources and also proposes an Extensible Markup Language (XML) based syntax based on this model [22]. Metadata attributes and semantics are defined as a RDF Schema which provides information about how RDF statements are to be interpreted. RDF aims to facilitate modular interoperability among different metadata element sets by creating what Eric Miller of OCLC calls "an infrastructure that will support the combination of distributed attribute registries" or complementary, independently-maintained metadata packages [23].

RDF has potential applications in a number of specific areas on the Web. It can be used to describe the content of resources, to model relationships between resources and to enable collections of Web pages to be represented as a single logical "document" - all of which can aid the resource discovery process. RDF also has potential applications for implementing content ratings services, for describing intellectual property rights and for dealing with privacy issues. The *RDF Schema Specification* additionally points out that "RDF with digital signatures will be key to building the 'Web of Trust' for electronic commerce, collaboration, and other applications" [24].

Many persons involved in the development of Dublin Core have been actively involved in the development of RDF, which can be understood in Dublin Core terms as an implementation of the conceptual Warwick Framework. A draft RDF schema for Dublin Core has been produced as part of the development of the *RDF Schema Specification* and it is possible that Dublin Core could be one of the earliest RDF implementations on the Web [25]. The twin development of Dublin Core and RDF has also resulted in a review of the data model that underlies Dublin Core that will feed into future developments of the format.

3 Integrating access to distributed and heterogeneous information objects

The development of core metadata standards and Web-based infrastructures will not immediately lead to the integration of resource discovery in what is an extremely heterogeneous environment. Historically, different types of repositories have developed their own ways of describing the content of their collections. Libraries have developed standards for describing library-based materials including images, using cataloguing rules like AACR2. This information could then be encoded in one of the Machine Readable Cataloguing (MARC) formats. Archives use a different family of standards including ISAD(G) - the *General International Standard Archival Description.* Museums and art galleries, if anything, are even more diverse. Individual museum departments might document their own artefacts in a number of different ways depending upon the provenance and format of items. Jim Blackaby and Beth Sandore have commented that for "a variety of reasons - availability of software, variations in standards, the needs of collections - heterogeneity can be expected to persist and in many cases, it should be encouraged" [26]. When other organisations that might hold relevant collections are added - photographic libraries, zoological and botanical gardens, local history societies, individuals, etc. - one gets an idea of the potential diversity and heterogeneity of image-based data and its associated metadata. This necessary heterogeneity is also replicated in the digital environment. User expectations, however, mean that there is a need to work towards integrating the resource discovery process in this diverse environment. Some of these issues have been considered by the UK MODELS project and by the development of a MODELS Information Architecture (MIA) for broker services [27].

3.1 Consortium for the Computer Interchange of Museum Information

Several important metadata initiatives with direct relevance to these issues have originated in the cultural heritage sector, primarily from the museum community. Investigation into standards for the interchange of cultural heritage information has been co-ordinated by the Consortium for the Computer Interchange of Museum Information - the CIMI Consortium. This consortium aims to provide demonstrations of how distributed and heterogeneous data can be accessed in a consistent way. One of CIMI's first projects was CHIO (Cultural Heritage Information Online). This project offered a demonstrator that explored the use of SGML (CHIO Structure) and the utility of the Z39.50 protocol (CHIO Access) [28, 29]. The project resulted in the publication of a CIMI Z39.50

Application Profile for cultural heritage information [30]. Current projects include a Dublin Core testbed that will explore the use of DC qualifiers, the emerging DC data model and the use of RDF.

3.2 Museum Educational Site Licensing Project

The Museum Educational Site Licensing Project (MESL) was a collaboration between museums and educational institutions in the United States, with support from the Getty Art History Information Program and MUSE Educational Media, to investigate the capture, distribution and educational use of digital images and their corresponding text descriptions [31]. The project acknowledged the uncertainty that exists with regard to intellectual property rights - particularly for image-based material - and attempted to bring stakeholders together to explore the technical, administrative and legal issues surrounding the educational use of networked digital images. As part of this, MESL developed a means of searching across diverse metadata by mapping these to a data dictionary with thirty-two fields [32].

3.3 Aquarelle

Projects based in Europe have looked at similar issues. The Aquarelle project, funded by the European Union (EU) under its Telematics Applications Programme, investigated technical and architectural solutions to the problem of linking distributed and diverse cultural heritage information sources. In co-operation with the CIMI Consortium, the Aquarelle Consortium developed and designed an architecture for an integrated resource discovery system based on Z39.50, producing a new Z39.50 Application Profile [33, 34]. The project also created an environment for the creation, dissemination and maintenance of documents called ‘folders’ that contain secondary information about cultural heritage resources [35].

3.4 The Electronic Library Image Service for Europe

Another EU funded project investigating access to distributed and heterogeneous information objects, this time with specific reference to digital images, is the Electronic Library Image Service for Europe (ELISE). This project was primarily concerned with building a complete digital image service that includes support for user registration and validation, rights management and charging mechanisms as well as being able to search across heterogeneous databases, again using Z39.50 [36, 37].

3.5 The Arts and Humanities Data Service

A service with similar aims is the UK Arts and Humanities Data Service (AHDS). The AHDS is funded by the Joint Information Systems Committee of the UK's Higher Education Funding Councils to collect, describe, and preserve the electronic resources which result from research and teaching in the humanities. It consists of five distributed subject-based service providers that give access to a variety of digital information in the subject domains of archaeology, history, literary, linguistic and other textual studies, the visual arts, and the performing arts. All service providers need to give access to catalogues of their own metadata and will need to operate within a resource description context specific to their own subject domain. For example, the Oxford Text Archive - the AHDS service provider for literary and linguistic texts - would normally describe resources using Text Encoding Initiative (TEI) headers - an implementation of SGML [38, 39].

Two of the other providers have a specific interest in image-based information. The Visual Arts Data Service collects digital resources of interest to the visual arts community [40]. The Performing Arts Data Service covers the broad fields of the performing arts including film, theatre, music, dance and the broadcast arts [41]. However at a higher level, the AHDS needs to implement a resource discovery system that will provide unified access to the resource description systems of all service providers.

The AHDS, in conjunction with UKOLN, organised a number of metadata workshops to discuss resource description issues for each of the subject domains covered by the service providers [42]. The eventual technical solution adopted a 'layered' approach to cross-domain resource discovery as outlined in the MODELS workshop series. At the highest layer, a layered system utilises a simple generic metadata format for basic resource discovery. At lower layers, the same system can be configured to use descriptive information from domain-specific metadata formats. Rosemary Russell characterises this as enabling a user, "in a single search environment, to 'drill down' or move progressively through a hierarchy of increasingly rich and specialist metadata as they ... [move] through a continuum from resource discovery to resource evaluation, access, and use" [43]. The AHDS system has been implemented using Dublin Core and a Z39.50 gateway [44].

4 Further applications of metadata

4.1 Models of the research process

It is becoming increasingly realised that metadata has potential applications in areas other than resource discovery. David Bearman and Jennifer Trant, for example, have modelled the different strands of the humanities research process and noted the essential role of metadata within the model [45]. This model proposes a five-stage research process:

- Discovery
- Retrieval
- Collation
- Analysis
- Re-presentation

In the electronic environment, each one of these stages will depend upon the creation and maintenance of appropriate metadata. The initial discovery stage is almost completely dependent upon the existence of metadata. This data should allow a researcher to make a decision about a resource's likely relevance and whether to request its retrieval. The retrieval stage may require further metadata. In a Web context it may just require the knowledge of an Uniform Resource Locator (URL) or other unique identifier. However, in other environments, there may be a need for technical data about file formats, dependencies on particular pieces of hardware and software and other information necessary to know whether it is possible to retrieve the resource in a meaningful way. Additionally, metadata about rights management - the terms and conditions under which a resource can be used - will also be important. Once a resource has been retrieved, it moves into the 'space' of the user for collation, analysis and re-presentation. The re-presentation of information will result in the creation of new resources which may require further reference to rights management metadata and will also require the creation of further metadata to document the new resource and the processes that led to its creation.

4.2 Image standards

Howard Besser of the University of Michigan has proposed several distinct areas where metadata standards need to be developed for digital images. He has said that the library and information community - amongst others - need to take "the steps necessary to ensure that digital images produced today will be viewable well into the future, and a key step in making that happen is the provision of adequate metadata" [46]. This metadata comes in six broad categories [47].

- The technical information required to view the image. This metadata would comprise information about image types - whether bitmaps, vector files or video - with information on particular file formats (e.g. TIFF, GIF, etc.), compression (e.g. JPEG) and colour metrics.
- Information about the image capture process. This would include metadata about the type of image digitised with information about the size and dimensions of the original object. Technical information about the maker and model of scanner hardware used together with details of what has been done to each image would also be useful.
- Information about the quality and veracity of an image. Users of images may need to know who was responsible for its digitisation. For example, a digital image created by a museum or art gallery from scanning the original or a high-quality surrogate may need to be differentiated from the same image scanned by a private individual at home.
- Information about the original object. It will in many cases to important to know the precise nature and origin of the source object and it link with any surrogates. This descriptive information may include legacy content metadata, including subject classifications.
- Information about an image's authenticity. One of the problems with all digital information is that it is difficult to be sure that an information object is what it claims to be. This is what Peter Graham calls 'intellectual preservation' [48]. How will users know that the digital object that they retrieve is the one that they want? How will administrators of digital repositories know that their holdings have not been subject to unauthorised changes, either accidental or deliberate? Solutions to these problems are likely to depend upon cryptographic techniques or implementations of digital signatures.

- Information about rights management. Rights information is basic to the use and reuse of image resources. Rights metadata might include information on viewing and reproduction restrictions and contact information for the rights holders. Publishers and other rights owners are beginning to investigate the potential of metadata for rights management [49].

Howard Besser suggests that standards need to be developed for all of these metadata types. Decisions also need to be made concerning the location of this metadata. Some information, for example core resource discovery data or rights management metadata, might best exist in an image header while other information might more suitably be stored in a separate (but linked) database. It is essential that much of this information is recorded (or captured) at the time when a digital image is created.

4.3 The Making of America II Testbed Project

Development of standards in these areas is under progress. The Making of America II (MoA II) testbed project – a US project investigating the creation of an integrated (but distributed) digital library of surrogates of archival material - has defined three distinct types of metadata which can aid the discovery, navigation and administration of resources.

- Descriptive metadata, primarily for resource discovery.
- Structural metadata, which are those metadata relevant to the representation of the digital object to the user.
- Administrative metadata, defined as the information necessary for a repository to manage its digital collections.

The MoA II White Paper identifies many of the metadata types described by Besser as administrative metadata and specifies a list of metadata fields (or elements) [50].

5 Digital preservation

One other potential application of metadata is in the context of the long-term preservation of digital information objects. The issues that surround digital preservation have been receiving increased attention in recent years and the problems are seen to be as much strategic as technical [51, 52]. It is becoming increasingly apparent that successful preservation strategies will depend upon the creation and maintenance of relevant documentation or metadata [53, 54].

For example, Stephen Robertson has developed a model of a Digital Rosetta Stone (DRS) - a detailed 'metaknowledge archive' that could store "the vast amounts of knowledge needed to recover digital data from ... superseded media and to reconstruct digital documents from their original formats" [55, 56]. Others have coined phrases like 'super-metadata' or 'digital tablets' to describe much the same concepts [57, 58]. A number of projects and initiatives exist which are attempting to identify the metadata types necessary for long-term digital preservation.

5.1 The RLG Working Group on Preservation Issues of Metadata

In 1997, the Research Libraries Group constituted a Working Group on the Preservation Issues of Metadata and its final report [59] is a good assessment of the preservation and metadata requirements of digital imaging technology. The working group limited itself to a consideration of the data elements that describe digital image files, arguing that other specialist groups could be constituted to analyse other formats when the need becomes more pressing. The metadata elements deemed crucial for the continued viability of a digital master file are outlined in Figure 3.

Element	**Brief description**
Date	Date file is created
Transcriber	Name of agency (or individual) responsible for transcribing the metadata
Producer	Agency (or individual) responsible for the physical creation of the file.
Capture Device	Make and model of digital camera or scanner.
Capture Details	Name of scanner software, version information, scanner settings, gamma correction, etc. Digital camera lens type, focal length, light source type, etc.

Change History	A record of modifications made to the file.
Validation Key	A mechanism allowing one to verify that the electronically transmitted file is what it purports to be.
Encryption	The technique by which data is encrypted before transmission.
Watermark	Indicates whether (or not) some bits in the file have been altered in order to create a digital fingerprint or similar.
Resolution	Resolution determined by pixel dimensions, pixels per inch or dots per inch.
Compression	Indicate whether (or not) file has been conversed.
Source	Physical characteristics of the source, etc.
Color	Pixel depth.
Color Management	Systems (if any) used to improve consistency of colour.
Color Bar/Gray Scale Bar	Indicates presence (or not) of either, with type.
Control Targets	Information about targets included in scanned file.

Figure 3: RLG Preservation Metadata elements. Source: RLG Working Group (1998)

The RLG Working Group also examined two 'core' metadata formats, the Dublin Core and the Program for Co-operative Cataloging's USMARC-based core record standard, so that the group could specify the metadata elements extra to these core element lists that would be important to serve preservation needs. The report also published examples of preservation metadata implemented in extended Dublin Core, in an USMARC record and in a simple XML-based format.

5.2 Reference Model for an Open Archival Information System

One of the most influential initiatives in the digital preservation area is the high-level model for a digital archive known as the *Reference Model for an Open Archival Information System* (OAIS) published by the Consultative Committee for Space Data Systems (CCSDS). OAIS is an ISO initiative (co-ordinated by the CCSDS) that defines a reference model for organisations concerned with the long-term preservation of digital information obtained from observations of terrestrial and space environments, but which should be applicable to other long-term digital archives [60]. The OAIS model has a 'taxonomy of archival information object classes' that includes:

- Content information. The information that is the primary object of preservation. This contains the primary Digital Object and any Representation Information (RI) needed to transform this object into meaningful information.
- Preservation Description Information (PDI). Any information necessary to adequately preserve the Content Information with which it is associated. It includes:
  - Reference Information, e.g. unique identifiers.
  - Context Information which may contain details on the subject of the information or its relationships with other objects.
  - Provenance Information giving details of publishers (or other rights owners) and documenting rights management issues.
  - Fixity Information that documents authentication mechanisms in use.
- Packaging Information. The information that binds and relates the components of a package into an identifiable entity on a specific media.
- Descriptive Information. The information that allows the creation of Access Aids - to help locate, analyse, retrieve or order information from an OAIS.

The OAIS reference model is an important attempt to develop consensus about how repositories can begin to address the long-term preservation of digital information objects.

5.3 National Library of Australia PANDORA logical data model

The National Library of Australia (NLA) developed a 'logical data model' for metadata in its PANDORA (Preserving and Accessing Networked DOcumentary Resources of Australia) digital preservation pilot project. This model is based on an entity-relationship diagram that identifies the logical entities that need to be supported by the PANDORA system. The highest level entities are:

- Identification
- Selection and negotiation
- Capture
- Preservation
- Rights Management and Access Control

Each of these is divided into further entities and each of these into metadata attributes. Preservation metadata is defined as "entities required to support the management of copies within the archive, including activities to ensure both the immediate and long term accessibility of the item". The entities include 'File' and 'File Type' (e.g. HTML, ASCII, JPEG, PDF, TIFF, etc.), 'Format' and 'Format Type' (e.g. Online, Diskette, CD-ROM, etc.). The notes on 'Format' suggest that such information should be recorded at the selection stage as part of a technical assessment. It also recommends that "a history trail is kept of the format of the copy at the time of archiving and any technical processing that has been conducted on the copy to ensure preservation and access" [61]. A copy of a publication may be converted from one format to another to improve accessibility in the host environment or to help the migration of whole categories of publication to a new technology base. Generally a conversion from one format to another will involve tangible formats (e.g., to transfer files from diskette to CD-R) but there may also be a requirement to convert data from a tangible to online format or vice versa. When a format is converted to another format type, a record will be maintained of the conversion process, with a link to the new format type.

The NLA is currently in the process of setting up a Digital Services Project (DSP) that will take on many of the functions of the PANDORA pilot but which will be more fully integrated into the library's other activities [62].

5.4 CURL Exemplars in Digital Archives project (Cedars)

Cedars (CURL Exemplars in Digital Archives) is a UK project funded by the Electronic Libraries (eLib) programme and managed by the Consortium of University Research Libraries (CURL) [63]. The project has three complementary aims:

- To promote awareness about the importance of digital preservation, both amongst research libraries and their users, and amongst the data creating and data supplying communities upon which they depend.
- To identify and disseminate appropriate strategies so that individual libraries can develop collection management policies for digital information objects that are appropriate to their needs.
- To investigate and promote appropriate methods for the long-term preservation of different classes of digital resources typically included in library collections.

The Cedars project has recognised the importance of metadata creation and maintenance in ensuring the continuing viability of digital information objects. The Cedars Access Issues Working Group has produced a preliminary study of preservation metadata and the issues that surround it [64, 65]. This study will be used as a basis for the development of a preservation metadata implementation within the project.

6 Conclusions

This paper has attempted to show that metadata can be useful for a variety of applications with regard to image-based data. Firstly, despite the ongoing development and growing sophistication of content-based image retrieval techniques, metadata will remain useful for retrieval based on the known content and context of images. If standard formats (Dublin Core, MARC, etc.) or protocols (Z39.50) are utilised, retrieval of images can also be easily integrated with the retrieval of other digital - and non-digital - information objects. In addition, metadata can be utilised to carry out a range of other functions: representation, authentication, rights management and digital preservation. Metadata implementations like RDF/XML offer the means to define metadata schemata to

help manage these functions. The main difficulty with metadata is that it can be extremely time-consuming and expensive to create and maintain. The advantage is that when combined with content-based approaches, metadata can enhance information retrieval and fulfil a host of additional, but important, administrative roles.

7 References

1. Ferguson ES. Engineering and the mind's eye. MIT Press, Cambridge, Mass., 1992, p. 75.

2. Evans H. Practical picture research: a guide to current practice, procedure, techniques and resources. Blueprint, London, 1992.

3. Enser, PGB. Progress in documentation: pictorial information retrieval. Journal of Documentation 1995; 51: 126-170.

4. Heaney M. Filming a Bodleian manuscript. Bodleian Library Record 1983; 11: 105-109.

5. Gudivada VN, Raghavan VV. Content based image retrieval systems. IEEE Computer 1995; 28(9): 18-22.

6. Eakins JP. Techniques for image retrieval. Library Information Technology Centre, South Bank University, London, 1998 (Library & Information Briefings, 85).

7. Heery R, Powell A, Day M. Metadata. Library Information Technology Centre, South Bank University, London, 1997 (Library & Information Briefings, 75).

8. Dempsey L, Heery R. Metadata: a current view of practice and issues. Journal of Documentation 1998; 54 (2): 145-172.

9. Couprie LD. ICONCLASS. Art Libraries Journal 1983; 8(2): 32-49.

10. Petersen T. Art & Architecture Thesaurus, 2nd ed. Oxford University Press on behalf of the Getty Art History Information Program, New York, 1994.

11. Getty Information Institute, College Art Association. Categories for the Description of Works of Art, v. 1.0. Getty Information Institute, Los Angeles, Calif.,1996.
<http://www.gii.getty.edu/cdwa/>

12. Weibel SL, Lagoze C. An element set to support resource discovery: the state of the Dublin Core, January 1997. International Journal on Digital Libraries 1997; 1: 176-186.

13. Weibel S, Godby J, Miller E, Daniel R. OCLC/NCSA Metadata Workshop report. Online Computer Library Center, Dublin, Ohio, 1995.
Available: <http://purl.org/metadata/dublin_core_report>

14. Weibel S. Dublin Core: a simple content description model for electronic resources. NFAIS Newsletter 1998; 40(7): 1-3.

15. Dempsey L, Weibel SL. The Warwick Metadata Workshop: a framework for the deployment of resource description. D-Lib Magazine, July/August 1996.
Available: <http://www.dlib.org/dlib/july96/07weibel.html>

16. Lagoze C, Lynch CA, Daniel R. The Warwick Framework: a container architecture for aggregating sets of metadata. Cornell University, Ithaca, N.Y., 1996 (Cornell Computer Science Technical Report TR96-1593).
Available: <http://cs-tr.cs.cornell.edu:80/Dienst/UI/1.0/Display/ncstrl.cornell/TR96-1593/>

17. Weibel S, Miller E. Image description on the Internet: a summary of the CNI/OCLC Image Metadata Workshop. D-Lib Magazine, January 1997.
Available: <http://www.dlib.org/dlib/january97/oclc/01weibel.html>

18. Weibel SL, Lagoze C. An element set to support resource discovery: the state of the Dublin Core, January 1997. International Journal on Digital Libraries 1997; 1: 176-186

19. Weibel S, Iannella R, Cathro W. The 4th Dublin Core Metadata Workshop Report. D-Lib Magazine, June 1997.
Available: <http://www.dlib.org/dlib/june97/metadata/06weibel.html>

20. Weibel S, Kunze J, Lagoze C, Wolf, M. Dublin Core metadata for resource discovery. RFC 2413, 1998.
Available: <http://info.internet.isi.edu:80/in-notes/rfc/files/rfc2413.txt>

21. Hakala J, Hansen P, Husby O, Koch T, Thorborg S. The Nordic Metadata Project: final report. Helsinki University Library, Helsinki, 1998.
Available: <http://linnea.helsinki.fi/meta/nmfinal.htm>

22. Lassila O, Swick R (eds.). Resource Description Framework (RDF) model and syntax specification. World Wide Web Consortium, Cambridge Mass., 1999 (W3C Recommendation).
Available: <http://www.w3.org/TR/REC-rdf-syntax/>

23. Miller E. An introduction to the Resource Description Framework. D-Lib Magazine May 1998.
Available: <http://www.dlib.org/dlib/may98/miller/05miller.html>

24. Brickley D, Guha RV (eds.). Resource Description Framework (RDF) schema specification. World Wide Web Consortium, Cambridge Mass., 1999 (W3C Proposed Recommendation).
Available: <http://www.w3.org/TR/PR-rdf-schema/>

25. Weibel S, Hakala J. DC-5: The Helsinki Metadata Workshop: a report on the workshop and subsequent developments. D-Lib Magazine, February 1998.
Available: <http://www.dlib.org/dlib/february98/02weibel.html>

26. Blackaby J, Sandore B. Building integrated museum information retrieval systems: practical approaches to data organisation and access. Archives and Museum Informatics 1997; 11: 117-146

27. Dempsey L, Russell R, Murray R. A utopian place of criticism? Brokering access to network information. Journal of Documentation 1999; 55(1): 33-70.

28. Moen WE. Distributed access to cultural heritage information. Library Technology 1998; 3(4): 63-64.
Available: <http://www.sbu.ac.uk/litc/lt/1998/news955.html>

29. National Information Standards Organization. Z39.50: Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. NISO, Bethesda, Md., 1995.
Available: <http://lcweb.loc.gov/z3950/agency/>

30. CIMI Z39.50 Working Group. The CIMI Profile: Z39.50 Application Profile for Cultural Heritage Information, release 1.0. University of North Texas, School of Library and Information Sciences, Denton, Tex., 1998.
Available: <http://www.cimi.org/downloads/ProfileFinalMar98/cimiprofile1.htm>

31. Trant J. Exploring new models for administering intellectual property: the Museum Educational Site Licensing Project. In: Heidorn PB, Sandore B (eds.). Digital image access and retrieval. Graduate School of Library and Information Science, University of Illinois at Urbana- Champaign, Urbana-Champaign, Ill., 1997, pp. 29-41.

32. Blackaby J, Sandore B. Building integrated museum information retrieval systems: practical approaches to data organisation and access. Archives and Museum Informatics 1997; 11: 117-146.

33. Michard A (ed.). Final report: IE-2005 Aquarelle: sharing cultural heritage through multimedia telematics. INRIA, Le Chesnay, 1998.
Available: <http:aqua.inria.fr/Aquarelle/Public/EN/final-report.html>

34. Michard A, Christophides V, Scholl M, Stapleton M, Sutcliffe D, Vercoustre A-M. The Aquarelle resource discovery system. Computer Networks and ISDN Systems 1998; 30(13): 1185-1200.

35. Doerr M, Fundulaki I, Christophidis V. The specialist seeks expert views: managing digital folders in the AQUARELLE project. In: Museums and the Web 97. Archives & Museum Informatics, Pittsburgh, Pa., 1997.
Available: <http://www.archimuse.com/mw97/speak/doerr.htm>

36. Eyre J. Architecture for online museums of the future: a object server for the future (ELISE II). In: Museums and the Web 97. Archives & Museum Informatics, Pittsburgh, Pa., 1997.
Available: <http://www.archimuse.com/mw97/speak/eyre.htm>

37. Eyre J. Distributed image services. VINE 1998; 107: 65-72

38. Giordano R. The documentation of electronic texts using Text Encoding Initiative headers: an introduction. Library Resources and Technical Services 1994; 38(4): 389-401.

39. Giordano, R. The TEI header and the documentation of electronic texts. Computers and the Humanities 1995; 29(1): 75-84.

40. Gill T, Grout C. Finding and preserving visual arts information on the Internet. Art Libraries Journal 1997; 22/3: 19-25.

41. Duffy C. An introduction to the Performing Arts Data Service. Literary and Linguistic Computing 1997; 12(4): 277-281.

42. Miller P, Greenstein, D (eds.). Discovering online resources across the humanities: a practical implementation of the Dublin Core. UKOLN, Bath, 1997.
Available: <http://ahds.ac.uk/public/metadata/discovery.html>

43. Russell R. UKOLN MODELS 4: evaluation of cross-domain resource discovery. In: Miller P, Greenstein, D (eds) Discovering online resources across the humanities: a practical implementation of the Dublin Core. UKOLN, Bath, 1997, pp. 18-21
Available: <http://ahds.ac.uk/public/metadata/disc_04.html#ukoln>

44. Dempsey L, Russell R, Murray R. The emergence of distributed library services: a European perspective. Journal of the American Society for Information Science 1998; 49: 942-951

45. Bearman D, Trant J. Unifying our cultural memory: could electronic environments bridge the historical accidents that fragment cultural collections? In: Dempsey L, Criddle S, Hestletine, R (eds.). Information landscapes for a learning society. Library Association, London, 1999 (forthcoming)

46. Besser H. Image databases: the first decade, the present, and the future. In: Heidorn PB, Sandore B (eds.). Digital image access and retrieval. Graduate School of Library and Information Science, University of Illinois at Urbana- Champaign, Urbana-Champaign, Ill., 1997, pp. 11-28.

47. Besser H, Trant J. Describing image files: the need for a technical standard. Coalition for Networked Information, Fall Meeting, Orlando, Florida, 30 November 1994.
Available: <http://www.lib.virginia.edu/cataloguing/ala/getty.html>

48. Graham, P.S.,1994, Long-term intellectual preservation. In: Elkington N.E. (ed.). Digital imaging technology for preservation. Research Libraries Group, Mountain View, Calif., pp. 41-57.

49. Rust G. Metadata: the right approach. An integrated model for descriptive and rights metadata in e-commerce. D-Lib Magazine, July/August 1998.
Available: <http://www.dlib.org/dlib/july98/rust/07rust.html>

50. Making of America II. The Making of America II testbed project white paper. Version 2.0, September 15 1998.
Available: <http://sunsite.berkeley.edu/MOA2/>

51. Garrett J, Waters D. Preserving digital information: report of the Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access and the Research Libraries Group. Commission on Preservation and Access, Washington, D.C., 1996.
Available: <http://www.rlg.org/ArchTF/>

52. Beagrie N, Greenstein, D. A Strategic Policy Framework for Creating and Preserving Digital Collections. Arts and Humanities Data Service, London, 1998.
Available: <http://ahds.ac.uk/manage/framework.htm>

53. Rothenberg J. Metadata to support data quality and longevity. Proceedings of the 1st IEEE Metadata Conference, NOAA Complex, Silver Spring, Md., 16-18 April 1996.
Available: <http://www.computer.org/conferen/meta96/rothenberg_paper/ieee.data-quality.html>

54. Rothenberg J. Avoiding technological quicksand: finding a viable technical foundation for digital preservation. Council on Library and Information Resources, Washington, D.C., 1999.

55. Robertson SB. Digital Rosetta Stone: a conceptual model for maintaining long-term access to digital documents. Thesis (MSc), Air Force Institute of Technology, Graduate School of Logistics and Acquisition Management, Dayton, Ohio, 1996.
Available: <http://www.au.af.mil/au/database/research/ay1996/afit_la/rober_sb.htm>

56. Heminger A, Robertson S. Digital Rosetta Stone: a conceptual model for maintaining long-term access to digital documents. In: Sixth DELOS Workshop: Preservation of Digital Information, Tomar, Portugal, 17-19 June 1998. European Research Consortium for Informatics and Mathematics, Le Chesnay, 1998, pp. 53-58 (ERCIM-98-W003).
Available: <http://www.ercim.org/publication/ws-proceedings/DELOS6/>

57. Chilvers A, Feather J. The management of digital data: a metadata approach. Electronic Library 1998; 16(6): 365-372.

58. Kranch DA. Beyond migration: preserving electronic documents with digital tablets. Information Technology and Libraries 1998; 17: 138-148.

59. RLG Working Group on Preservation Issues of Metadata. Final report. Research Libraries Group, Mountain View, Calif., 1998.
Available: <http://www.rlg.org/preserv/presmeta.html>

60. Reich L, Sawyer D (eds.). Reference Model for an Open Archival Information System (OAIS). Consultative Committee for Space Data Systems, White Book, Issue 4. CCSDS Secretariat, National Aeronautics and Space Administration, Washington, D.C., 1998. (CCSDS 650.0-W-4.0).
Available: <http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html>

61. National Library of Australia. PANDORA Logical Data Model, version 2. National Library of Australia, Canberra, 1997.
Available: <http://www.nla.gov.au/pandora/ldmv2.html>

62. National Library of Australia. Digital Services Project. National Library of Australia, Canberra, 1999.
Available: <http://www.nla.gov.au/dsp/>

63. Day M. CEDARS, digital preservation and metadata. In: Sixth DELOS Workshop: Preservation of Digital Information, Tomar, Portugal, 17-19 June 1998. European Research Consortium for Informatics and Mathematics, Le Chesnay, 1998, pp. 53-58 (ERCIM-98-W003).
Available: <http://www.ercim.org/publication/ws-proceedings/DELOS6/>

64. Day M. Metadata for Preservation. UKOLN, Bath, 1998 (CEDARS Project Document AIW01).
Available: <http://www.ukoln.ac.uk/metadata/cedars/AIW01.html>

65. Day M. Issues and Approaches to Preservation Metadata. Joint RLG and NPO Preservation Conference: Guidelines for Digital Imaging, University of Warwick, 28-30 September 1998.
Available: <http://www.thames.rlg.org/preserv/joint/day.html>

8 Acknowledgements

UKOLN is funded by the British Library Research and Innovation Centre (BLRIC), the Joint Information Systems Committee (JISC) of the UK higher education councils, as well as by project funding from several sources. UKOLN also receives support from the University of Bath, where it is based.